PMC17

RASCH ANALYSIS OF THE FATIGUE IMPACT SCALE

Meads DM[1], Hampson NE[1], Fisk JD[2], McKenna SP[1], Doward LC[1], Mayo KW[3]
[1]Galen Research, Manchester, Manchester, UK; [2]Dalhousie University, Halifax, Nova Scotia, Canada; [3]Novartis Pharma AG, Basel, Basel, Switzerland

OBJECTIVES: The Fatigue Impact Scale (FIS) is a 40-item multiple response questionnaire designed to assess fatigue, divided into cognitive, physical, and psychosocial functioning subscales. The purpose of the study was to improve the measure by removing differential item functioning (DIF) and reducing the number of items. **METHODS:** FIS data were available from 188 patients with multiple sclerosis (MS). These data were subjected to Rasch analysis (one-parameter logistic item response theory) using the RUMM programme. Fit to the Rasch model was examined via Chi2 statistics and assessments of DIF related to gender, age and MS type. **RESULTS:** FIS responses from the 188 MS patients were analysed (47/25% male; mean age 50.9, SD 10.5; 39.6% relapsing remitting, 36.9% primary progressive, 23.5% secondary progressive). Initial results showed that the 40-item FIS exhibited misfit and was not unidimensional. Several items exhibited DIF by age, gender or MS-type. For example, patients aged over 50 years scored significantly higher than patients aged 50 or younger (who had a similar level of fatigue) on the item "I have to rely more on others to help me". DIF by MS-type indicated that answers to certain items are significantly influenced by the disease stage of the patient. After the removal of 9 items that either misfit or exhibited age or gender DIF, the reduced FIS fit the Rasch model (Chi2 $p > 0.05$), providing a unidimensional fatigue scale. The threshold map suggests that for some items the response categories did not discriminate in the way intended, suggesting that changing the response format may improve the scale. **CONCLUSIONS:** The analysis showed that it is feasible to derive a single unidimensional scale of fatigue from the FIS. The severity (logit) coverage of the scale is good but there remains item redundancy suggesting further scope for item reduction.

PMC18

PSYCHOMETRIC PROPERTIES OF A TOUCH SCREEN COMPUTER-BASED VERSION OF THE SF-36

Ramachandran S[1], Taber T[2], Coons SJ[1]
[1]University of Arizona, Tucson, AZ, USA; [2]Assist Technologies, Scottsdale, AZ, USA

OBJECTIVES: The purpose of this study was to examine the psychometric properties of a touch screen version (Assist Technologies) of the SF-36, a widely used measure of self-reported health status. **METHODS:** Non-probability purposive sampling was used to recruit 300 subjects intended to reflect the primary socio-demographic characteristics of the US general adult population. The SF-36 was administered via touch screen along with the EQ-5D and other items. Amount of missing data and presence of floor and ceiling effects were assessed. Scale score internal consistency was estimated using Cronbach's alpha coefficient. As one test of construct validity, mean scale scores were compared across groups known to differ in regard to presence of chronic conditions. Convergent and discriminant construct validity were evaluated through examination of correlations between SF-36 scales and the EQ-5D domains. **RESULTS:** A total of 312 respondents completed the study. Scale means and standard deviations for the touch screen SF-36 in this sample were very similar to those seen with the paper-based format in the US general population. Less than 1% of all responses were missing. The percentage of respondents at the floor for almost all scales was less than 10%. Ceiling effects were evident for several of the scales. In general, these floor and ceiling effects were very similar to that observed in the general US population. All of the reliability coefficients exceeded 0.70; the range was from 0.75 to 0.93. Respondents with one or more chronic conditions reported significantly lower scores on all eight scales of the SF-36 compared to those with no chronic conditions. The direction and strength of the correlations between the SF-36 scales and the EQ-5D domains were as hypothesized. **CONCLUSION:** The comparable psychometric properties and lower levels of missing data make this touch screen version a very viable alternative to the paper-based SF-36 format.

PMC19

HOW LONG AGO . . . ?: ASSESSING PATIENT ADHERENCE TO SPECIFIED QUESTIONNAIRE RECALL PERIODS

Coyne KS, Brewster-Jordan J
The MEDTAP Institute at UBC, Bethesda, MD, USA

OBJECTIVES: Patient-reported outcome (PRO) measures are typically designed to instruct patients to consider a specific timeframe (recall period) when answering each item. Recall periods vary in length (i.e., 24 hours, 1 week, or 4 weeks) based upon the condition being assessed and the objectives of the research. The goal of this study was to assess patient adherence patterns to recall periods varying in duration by analyzing summaries of one-on-one qualitative patient interviews. **METHODS:** Data were reviewed from eight previously conducted cognitive debriefing interviews on condition specific measures that varied in recall period length. In all interviews, patients were specifically asked what recall period they had used when completing the PRO measure. The patient's response and the questionnaire's prespecified recall period were compared. **RESULTS:** Cognitive debriefing data for 115 patients (55% women) with a mean age of 57.1 was reviewed. The conditions of the ten PROs evaluated were: GERD (n = 2), Dementia (n = 2), Diabetes (n = 3), and Overactive Bladder (n = 3). Recall periods were: Daily (n = 1); 1 week (n = 2); 2 weeks (n = 2); 2 to 4 weeks (n = 1); and 4 weeks (n = 4). The majority of patients (57.9%) stated the recall period specified on the PRO measure; 14.5% recalled a general period of time (e.g. since they had the condition); 13.8% stated a time over the recall period while 12.6% stated a time under the recall period. Shorter recall periods (e.g. 1 week) had more concordant patient responses than longer recall periods (80% vs. 53%). **CONCLUSIONS:** Patients tend to adhere better to shorter recall periods than longer recall periods when completing PRO measures. Questionnaires with longer recall periods often result with patients thinking in general terms of their condition or using a recall period of their own.

PMC20

INDIRECT COMPARISON (OR COMMON-COMPARATOR) METHODS FOR META-ANALYSIS OF SUMMARY DATA

Fitzgerald P, LeReun C, Aristides M
M-TAG, A division of IMS Health Economics and Outcomes Research, London, UK

OBJECTIVES: In this presentation we summarise statistical methods for meta-analysis when a direct comparison between treatment effects is impossible, inadequate, or inappropriate. **METHODS:** Detailed descriptions are presented, and these are appraised per se and in relation with conventional meta-analysis methods. The main methods can be summarised as follows: weighted mean difference of relative effect measures (e.g. mean difference, log-odds-ratio, log-relative-risk and log-hazard-ratio) and meta-regression of relative effect measures, both of which are based on traditional meta-analysis approaches, and weighted

Bayesian regression models, which are more flexible and are simple to implement in freely available software. **RESULTS:** Using health outcomes research examples for illustration in each case, we describe common methodology issues arising from use of these methods, such as when small numbers of trials are analysed, when unequal trial sizes are included and when excess variability between trials (or heterogeneity) is encountered. **CONCLUSIONS:** For the methods considered, we offer possible solutions, make recommendations for their use and point out situations in which caution should be exercised.

PMC21

EVALUATING THE DIFFERENCE BETWEEN AVERAGE WHOLESALE PRICE AND WHOLESALE ACQUISITION COST FOR PHARMACEUTICALS IN THE UNITED STATES

Malone DC, Mahmood M

University of Arizona College of Pharmacy, Tucson, AZ, USA

OBJECTIVES: 1) To examine the percent difference between average wholesale price (AWP) and wholesale acquisition cost (WAC) for pharmaceuticals in the United States, accounting for patent status and manufacturer type, and 2) to evaluate the relationship between brand manufacturers and relabelers. **METHODS:** Data for this study came from the Master Drug Data Base (MDDB), which is a proprietary drug file containing pricing information for all prescription and non-prescription products available in the US. The percent difference between AWP and WAC for prescription pharmaceuticals was compared on a variety of facets, including single source, type of manufacturer (original or repackager). The difference was expressed as a percentage of AWP (a commonly used method for reimbursing pharmacies in the US). We also compared the AWP among brand name manufacturers and relabelers (who repackage brand name pharmaceuticals produced by the original manufacturer). **RESULTS:** A total of 23,607 unique drug products were included in the analysis examining AWP and WAC. The mean percent difference for brand name pharmaceuticals was 0.23 + 0.11, as compared to 0.44 + 0.26, $p < 0.001$. Brand name drugs that were available from multiple companies had a mean difference of 0.25 + 0.14, compared to 0.20 + 0.05 for single source products ($p < 0.001$). The median AWP for brand name manufacturers was $3.04 per unit, compared to $3.11 per unit for relablers. **CONCLUSION:** This study documents the magnitude of well-known differences between AWP and WAC for brand name and generic products. Further, branded products produced by more than one manufacturer will have larger differences between AWP and WAC than single source products. The findings suggest the need for analysts to critically evaluate the use of AWP for determining product costs in the US and substantial differences exist between single source and multiple source products. A more transparent and accurate pricing system is needed for economic analyses in the US.

PMC22

RELATIVE WEIGHTS ASSIGNED TO PHARMACY PROCEDURES: OPPS METHODS DESIGN AND CONCEPTS

Baker JJ

The Resource Group, Dallas, TX, USA

OBJECTIVE: The Centers for Medicare and Medicaid Services (CMS) assigns relative weights to procedures in the hospital outpatient setting and may do so for pharmacy handling costs in 2006. This study examines the current approach to assignment of relative weights for drug and biological administration codes under the CMS Hospital Outpatient Prospective Payment System (OPPS) and compares this approach to the resource-based level of effort concept initially created for payment to physicians. **METHODS:** Resource-based methods originally proposed for the hospital OPPS and equivalent measurement methods in the original Harvard RBRVS work were deconstructed and evaluated. The MedPAC rationale for recommendation of relative weights for payment of pharmacy handling costs and the proposed weights were also examined and evaluated. **RESULTS:** The evaluation sought indications of resource-based level of effort applications in the OPPS and their comparability to the original resource-based studies, especially in the area of intensity measures. The underlying intent of the resource-based relative value scale (RBRVS) was to create a hierarchy of resource-based level of effort in physician service delivery. The concept of hospital OPPS was also intended to reflect resource-based service delivery. Procedures are assigned a relative weight, implying that payment includes level of effort resources. Recommended 2006 handling costs for drugs are assigned five relative weights, each compared to drawing up an injected drug for administration. We postulate these relative weights contain insufficient recognition of the level of effort and resource consumption required, thus distorting the concept's initial intent. **CONCLUSIONS:** Many researchers and policy makers assume that relative weights equate to level of effort resource consumption in all instances. We cannot find this is so regarding the five levels of relative weights recommended for OPPS pharmacy handling costs. Because resource consumption is disproportionately greater than relative weights assigned in many cases, the resulting payment will be understated.

PMC23

ECONOMIC EVALUATION OF MEDICAL DEVICES IN FRANCE: A CHALLENGE FOR HEALTH ECONOMISTS

Späth HM[1], Taboulet F[2], Carrère MO[1]

[1]GRESAC, Centre Léon Bérard, Lyon, France; [2]University Toulouse 3, Toulouse, France

OBJECTIVES: Pharmacoeconomic evaluation activities have grown rapidly in recent years, but few economic evaluations have focused on Medical Devices (MD). This study addresses the barriers to conducting economic evaluations of MD, in comparison with pharmacoeconomic evaluations, in order to develop a framework for MD economic evaluation. **METHODS:** First, we studied the differences between MD and drugs that possibly impact on the completion of economic evaluations. Then, we analysed items of the French Guidelines for Economic Evaluations of Health Care Technologies developed by the "Collège des Economistes de la Santé" [http://www.ces-asso.org/docs/France_Guidelines_HE_Evaluation.PDF] that might be barriers to the completion of MD economic evaluations, as compared to drugs, and we developed suggestions to overcome these barriers. **RESULTS:** In this abstract we present three of eight barriers to performing economic evaluations of MD. The first one relates to the feasibility of clinical trials, e.g. inadequacy of "placebo" and "double blinding" for MD testing and difficulty to include large numbers of patients. We suggest performing comparative studies for assessing clinical outcomes to be included in economic evaluations and to discuss potential bias. Secondly, MD is developed by engineers who are used to assessing technical performance, but not clinical and economic outcomes. We propose setting up collaborations between engineers, health care professionals and health economists from the very beginning of MD development. Besides, MD effectiveness often depends on the operator (health care professional or patient) and may change over time, when the operator gains experience. Health economists must, therefore, analyse the transferability of economic evaluation results from one setting to another and over time. **CONCLUSION:** We recommend setting up multidisciplinary groups of engineers,